

# SCIENTIFIC REPORTS

OPEN

## Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability

Received: 01 March 2016

Accepted: 20 April 2016

Published: 06 May 2016

Carlo Camilloni<sup>1,2</sup>, Benedetta Maria Sala<sup>3</sup>, Pietro Sormanni<sup>1</sup>, Riccardo Porcari<sup>4</sup>, Alessandra Corazza<sup>5</sup>, Matteo De Rosa<sup>3</sup>, Stefano Zanini<sup>5</sup>, Alberto Barbiroli<sup>6</sup>, Gennaro Esposito<sup>5,7</sup>, Martino Bolognesi<sup>3,8</sup>, Vittorio Bellotti<sup>4</sup>, Michele Vendruscolo<sup>1</sup> & Stefano Ricagno<sup>3</sup>

A wide range of human diseases is associated with mutations that, destabilizing proteins native state, promote their aggregation. However, the mechanisms leading from folded to aggregated states are still incompletely understood. To investigate these mechanisms, we used a combination of NMR spectroscopy and molecular dynamics simulations to compare the native state dynamics of Beta-2 microglobulin ( $\beta 2m$ ), whose aggregation is associated with dialysis-related amyloidosis, and its aggregation-resistant mutant W60G. Our results indicate that W60G low aggregation propensity can be explained, beyond its higher stability, by an increased average protection of the aggregation-prone residues at its surface. To validate these findings, we designed  $\beta 2m$  variants that alter the aggregation-prone exposed surface of wild-type and W60G  $\beta 2m$  modifying their aggregation propensity. These results allowed us to pinpoint the role of dynamics in  $\beta 2m$  aggregation and to provide a new strategy to tune protein aggregation by modulating the exposure of aggregation-prone residues.

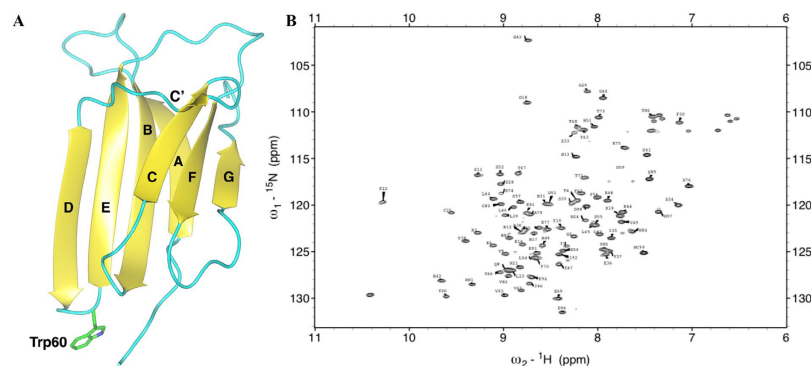
Protein misfolding and aggregation is a widespread phenomenon associated with the intrinsic properties of polypeptide chains<sup>1,2</sup>, indeed, it has been observed that the cross- $\beta$  structure characteristic of amyloid aggregates often represents the most thermodynamically stable state of polypeptide chains<sup>3,4</sup>. Thus, at least at typical cellular concentrations, proteins can be thermodynamically metastable in their native states and do not aggregate only because of the presence of high kinetic barriers<sup>3,4</sup>. It is therefore crucial to uncover the strategies that proteins adopt to remain soluble and escape aggregation *in vivo*, and shedding light on the design principles against aggregation would allow a better control of this process<sup>5,6</sup>.

Protein aggregation mechanisms can be identified from the analysis of the many amyloidogenic mutants that have been characterized in recent years, including those of transthyretin, lysozyme,  $\beta 2m$ , and gelsolin<sup>5,6</sup>. Known mutations in these proteins destabilize their native states, often by increasing their dynamics and flexibility<sup>7–11</sup>. When destabilized, native states can undergo larger structural fluctuations, or even unfolding, thus favoring the formation of non-native interactions and, ultimately, the deposition into amyloid aggregates<sup>11</sup>.

To determine the molecular origins of these processes, we compare here wild type  $\beta 2m$  and its W60G mutant, to investigate the correlation of thermodynamic stability and conformational flexibility *versus* aggregation propensity.  $\beta 2m$  is a single-domain protein, characterized by a seven-stranded  $\beta$ -sandwich fold typical of the immunoglobulin domain family;  $\beta$ -strands within the protein are named from A to G (Fig. 1A)<sup>12</sup>.  $\beta 2m$  is the light

<sup>1</sup>Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. <sup>2</sup>Department of Chemistry and Institute for Advanced Study, Technische Universität München, Lichtenbergstraße 4, D-85748 Garching, Germany.

<sup>3</sup>Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy. <sup>4</sup>Wolfson Drug Discovery Unit, Centre for Amyloidosis and Acute Phase Proteins, University College London, London NW3 2PF, UK. <sup>5</sup>Dipartimento di Scienze Mediche e Biologiche, Università di Udine, 33100 Udine, Italy. <sup>6</sup>Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente, Università degli Studi di Milano, 20133 Milano, Italy. <sup>7</sup>Science and Math Division, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE. <sup>8</sup>CIMAINA and CNR Istituto di Biofisica, c/o Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy. Correspondence and requests for materials should be addressed to C.C. (email: carlo.camilloni@ch.tum.de) or S.R. (email: stefano.ricagno@unimi.it)



**Figure 1. NMR assignment of  $\beta 2m$ .** (A) Ribbon representation of the crystal structure of human  $\beta 2m$  (pdb code 2YXF) with  $\beta$ -strands labeled according to standard nomenclature. W60 is shown as sticks. (B) HSQC [ $^1H$ ,  $^{15}N$ ] spectrum recorded at 11.7 T (500.13 MHz for  $^1H$ ), 310K, of [ $U$ - $^{13}C$ ,  $U$ - $^{15}N$  W60G  $\beta 2m$ ] 0.5 mM dissolved in 70 mM phosphate buffer at pH = 6.6 and 100 mM NaCl.

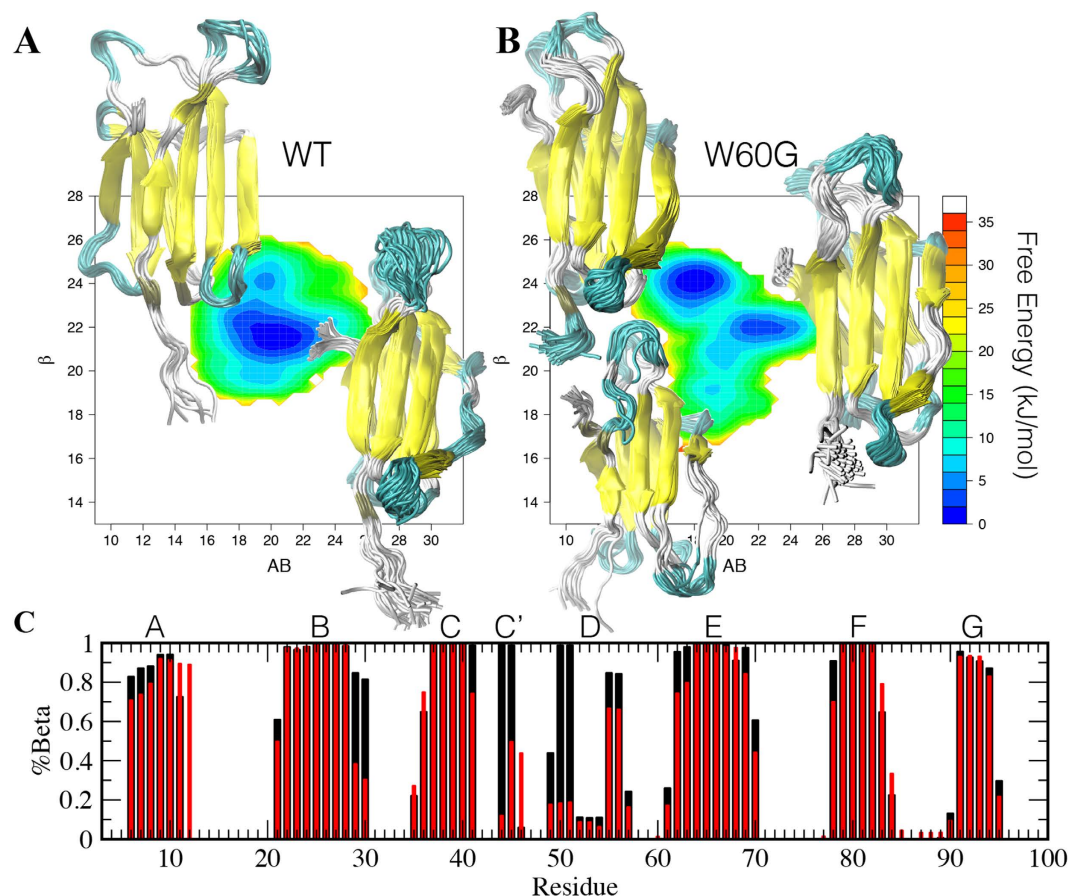
chain of the major histocompatibility complex (MHC) class I<sup>12–14</sup>. While  $\beta 2m$  is highly stabilized by the interactions within the MHC class I<sup>15</sup>, when  $\beta 2m$  is released as a monomer, it may turn into the etiological agent of dialysis-related amyloidosis, a condition triggered by renal impairment and the subsequent chronic accumulation of abnormally high concentrations of  $\beta 2m$  in body fluids.  $\beta 2m$  deposits in joints, bones and muscles, lead to movement impairment, bones fragility and articular pain<sup>16,17</sup>.

Recent efforts in elucidating the molecular determinants of  $\beta 2m$  aggregation have highlighted the complex role played by W60, an evolutionary-conserved, solvent-exposed residue. *In silico* simulations of wild-type  $\beta 2m$  suggested that W60 is heavily involved in intermolecular interactions<sup>18</sup>, therefore the non-natural mutation to Gly (W60G) was prepared, yielding to a variant with unexpected properties<sup>19</sup>. In particular, the (unfolding) thermodynamic stability of this mutant is higher than that of the wild-type  $\beta 2m$ , with  $\Delta G^\circ(H_2O)$   $22.2 \pm 2.0$  and  $27.6 \pm 3.3$  kJ mol<sup>-1</sup>, for wild-type and W60G variants, respectively<sup>19</sup>. While both wild-type and W60G aggregate under strongly denaturing conditions (pH 2.5), W60G shows low aggregation propensity using the standard protocol at pH 7.4 and 20% TFE, conditions under which wild-type  $\beta 2m$  aggregates abundantly<sup>19,20</sup>. A wealth of data indicates that the DE loop, where W60 is located, is under backbone geometrical strain in wild-type  $\beta 2m$ <sup>18,21–23</sup>. The introduction in the DE loop of a residue endowed with higher conformational freedom, such as Gly, relieves the local conformational strain and results in a  $\beta 2m$  variant thermodynamically more stable in solution, with a decreased aggregation propensity<sup>19,24</sup>. The above observations prompted us to select wild-type  $\beta 2m$  and its W60G variant as a well-suited system to investigate, besides protein thermodynamic stabilities, other mechanisms that may tune protein aggregation propensity.

To gain further understanding into the molecular bases of the different solution properties displayed by wild-type and W60G forms, we characterized the equilibrium distribution of conformations for the two variants. We employed NMR solution chemical shifts measured under native conditions in combination with replica-averaged metadynamics (RAM) simulations<sup>25–27</sup>. RAM simulations allow integrating the physico-chemical knowledge of a system (*i.e.* the interactions among its atoms as defined in molecular force fields) with the knowledge derived by experimental measurements (*i.e.* chemical shifts), to provide an atomistic description of the Boltzmann distribution of structures that satisfy the experimental data and the maximum entropy principle<sup>28</sup>. Comparison of the resulting ensembles shows how the W60G mutation decreases the overall aggregation propensity by increasing the average  $\beta$ -structure content and by perturbing the overall dynamics of surface residues. To test whether a better average protection of aggregation-prone surface residues can actually be key in regulating protein aggregation, we designed surface mutations aiming to either decrease the aggregation propensity of wild-type  $\beta 2m$ , or increase that of the W60G variant. We found that the designed mutants do not cause any relevant structural rearrangement in the protein core, nor they alter  $\beta 2m$  thermodynamic stability, but display striking different aggregation propensities. Thus, the approach presented here helps to decouple different molecular determinants contributing together to determine a high aggregation propensity, this strategy may be used to understand and modulate the aggregation properties of other folded proteins without affecting their structures and thermal stabilities.

## Results

**NMR chemical shift assignment of the W60G mutant.** Wild-type  $\beta 2m$  was already characterized in terms of  $^1H$ ,  $^{15}N$  and  $^{13}C$  backbone and side chain assignment<sup>29</sup>, whereas for the W60G only  $^1H$  resonances and amide nitrogens were previously attributed<sup>19</sup>. Here we extended the NMR analysis to the full backbone, as well as to  $C\beta$  and  $H\alpha$ , and in favorable cases to  $H\beta$ . The  $^{15}N$ - $^1H$  HSQC spectrum at 310 K (Fig. 1B), shows 108 (96 belonging to HNs) of the 110 peaks expected. The missing amide peaks belong to K58 and S88, which were not observed also in the wild-type type form due to unfavorable local exchange broadening<sup>30</sup>. In fact only at very low protein concentration (less than 50  $\mu M$ ), where the presence of dynamic dimers is scarce, these two peaks are visible. Moreover A15 and Y66 NHs overlap in the HSQC spectrum acquired at 500 MHz. The sequential backbone assignment was obtained using heteronuclear 3D triple-resonance experiments acquired on  $^{13}C$ ,  $^{15}N$



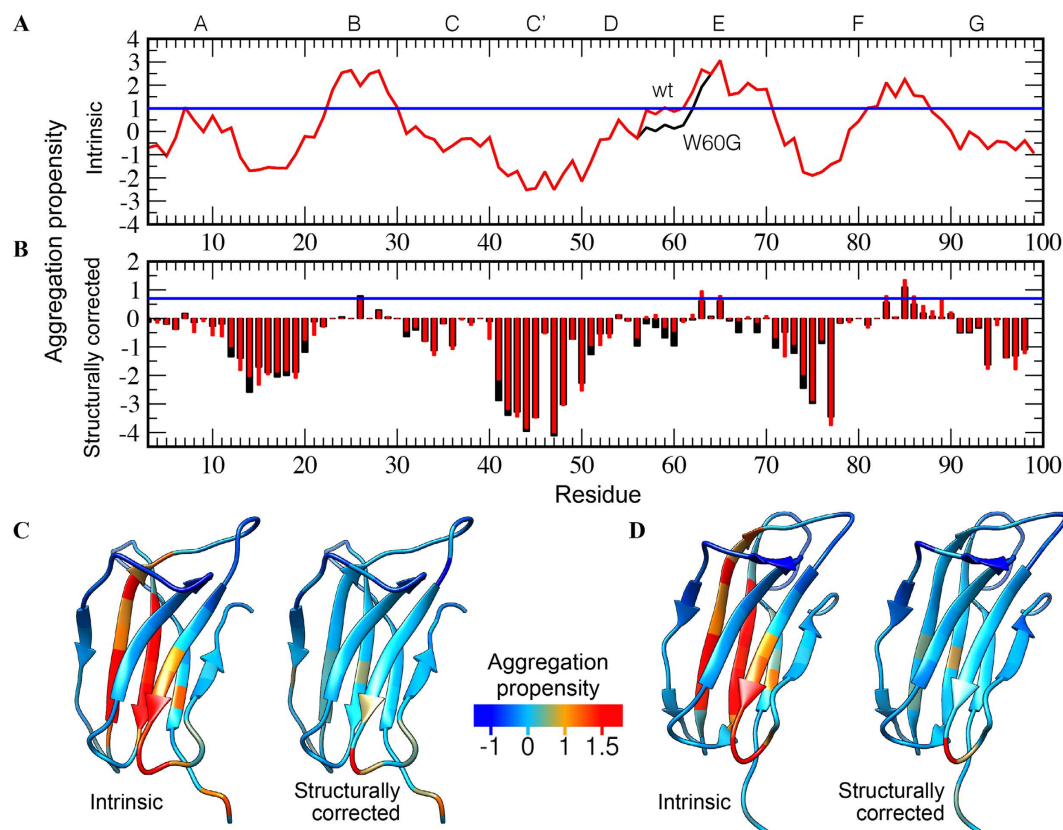
**Figure 2.**  $\beta$ 2m RAM Ensembles. Free energy surfaces (in kJ/mol) for wild-type (A) and W60G  $\beta$ 2m (B) as a function of the side chains rotameric state, AB, and the antiparallel  $\beta$ -structure content ( $\beta$ ). (C)  $\beta$ -structure populations for wild-type (red bars) and W60G (black bars)  $\beta$ 2m. The seven  $\beta$ -strands building the protein fold are identified by A through G labels.

uniformly labeled sample. The assignment percentage for C', C $\alpha$ , C $\beta$  were 93%, 97%, 86%, respectively. 95% of H $\alpha$  and only 23% of H $\beta$  were identified through (H)CCH-TOCSY spectra.

**Native state dynamics of  $\beta$ 2m.** We determined the conformational properties of wild-type and W60G  $\beta$ 2m using NMR chemical shifts and molecular simulations in the RAM framework<sup>25–27</sup>. In RAM the sampling of the conformational space is enhanced by employing bias-exchange metadynamics<sup>31</sup>, while the quality of the force field is improved by including the information content of equilibrium experiments, *i.e.* NMR chemical shifts<sup>25,26</sup>, within the framework of the maximum entropy principle<sup>28</sup>.

Using the RAM approach we obtained a converged sampling, resulting in a free energy landscape within a statistical uncertainty of less than 2 kJ/mol for free energies up to 20 kJ/mol (Figs S1 and S2). The free energy surfaces for the wild-type and W60G  $\beta$ 2m variants, as a function of the side-chain rotamer distribution (AlphaBeta collective variable, AB, see Methods) and the anti-parallel  $\beta$ -structure content (AntiBetaRMSD collective variable,  $\beta$ , see Methods), are shown in Fig. 2A,B. The free energy surfaces indicate that W60G hosts on average a higher content of  $\beta$ -structure than the wild-type protein. By closer inspection, the main differences in the  $\beta$ -sheet populations ( $\beta$ ) are localized in the B, C' and D strands (Fig. 2C). These differences show how the local relaxation of the strain of the backbone in position 60 propagates along the D strand up to C' due to its reduced freedom on one side and through space to the C-terminal of the B strand on the other. Wild-type  $\beta$ 2m shows a single well-defined minimum at  $\beta = 22$  and AB = 20, and a low-populated minimum at higher  $\beta$ -sheet content ( $\beta = 24$ ), compatible with a fully structured D strand. Conversely, W60G shows two minima of comparable energy, one at  $\beta = 24$  and AB = 18, and the other at  $\beta = 22$  and AB = 22; both minima are close to the two minima of the wild-type protein (Fig. 2). The effect of the W60G mutation on the dynamics of  $\beta$ 2m is also reflected by the reduced structural fluctuations in the 25–31 and 40–60 regions (B,C,C' and D strands) as shown in Fig. S3.

**Analysis of sequence and structural properties leading to protein aggregation.** Wild-type and W60G are characterized by significantly different aggregation properties. While the wild-type protein can form fibrils under both neutral (pH 7.4) and acidic (pH 2.5) conditions, the mutant yields fibrils only under denaturing (pH 2.5) conditions<sup>19,20</sup>. Such difference was explained in terms of the different thermodynamic stability of the two variants. NMR relaxation studies under conditions far from those favoring aggregation confirm that the

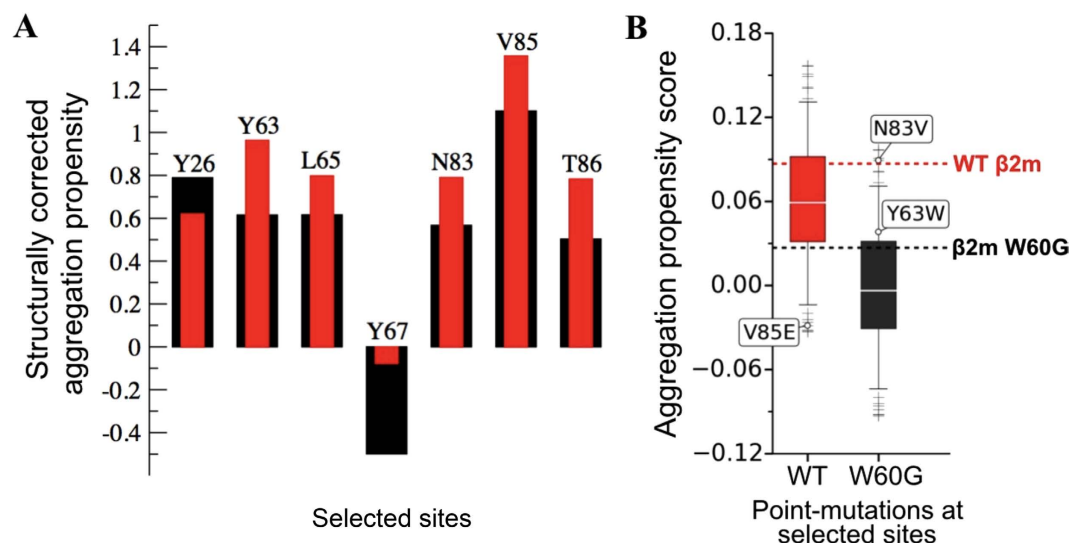


**Figure 3. Predicted aggregation properties of  $\beta 2m$  wild-type and W60G variants.** (A) Intrinsic aggregation profiles of the wild-type (red) and W60G (black) variants. (B) Structurally corrected profiles calculated as ensemble averages for  $\beta 2m$  wild-type (red bars) and W60G (black bars). In (A,B) values larger than 1 denote aggregation-prone regions. (C) Projection on the structure of wild-type  $\beta 2m$  of the intrinsic profile (left) and the structurally corrected profile (right). The color code is blue for aggregation-resistant residues and red for aggregation-prone residues. (D) Same as (C) for the structure of W60G  $\beta 2m$ .

wild-type species exhibits a much more pronounced association propensity than the W60G mutant<sup>30</sup>. However, additional insight can be obtained by analyzing the amyloidogenic properties of  $\beta 2m$  in terms of sequence- and structure-based properties. The intrinsic and structurally corrected aggregation propensities were assessed using the CamSol method (Fig. 3A,B)<sup>32</sup>, which is highly accurate in predicting mutation-induced changes in protein solubility, a property often strongly correlated with aggregation<sup>33</sup>. In the present work we used aggregation propensity scores that are the inverse of the CamSol solubility scores (so that higher scores represent higher aggregation propensity) and, in order to analyze the RAM ensemble, we have defined a ‘total structurally-corrected score’ (see Methods). The latter can be used to predict the tendency to aggregate of individual conformations of the ensemble; it is calculated from the structurally corrected CamSol profile in the same way the CamSol intrinsic score is calculated from the intrinsic profile<sup>32</sup>. The ensemble average of the total structurally corrected scores for the two  $\beta 2m$  variants is  $-0.59 \pm 0.01$  and  $-0.66 \pm 0.01$ , for the wild-type and W60G, respectively. Such scores are in keeping with the experimental aggregation data, *i.e.* wild-type  $\beta 2m$  is more aggregation prone than W60G  $\beta 2m$  under native conditions.

The intrinsic and the structurally corrected profiles are represented in Fig. 3C,D, color-coded on the structures of wild-type and W60G. Aggregation-prone residues are mostly located in the B, E strands, and at the C-terminal of the F strand (Fig. 3A). These  $\beta$ -strands are the central strands in the two  $\beta 2m$   $\beta$ -sheets; the B and E strands are flanked by the A and D strands (in the 4-stranded  $\beta$ -sheet), while the F strand is surrounded by the C and G strands (in the 3-stranded  $\beta$ -sheet) (Fig. 1A). It is noteworthy that all these flanking  $\beta$ -strands consistently prove less aggregation-prone than the central ones, according to both the intrinsic and the structurally corrected CamSol profiles (Fig. 3B), an observation that may be interpreted as a strategy to protect the protein from unwanted aggregation. Furthermore, at least half of the aggregation-prone residues point towards the protein core, thus being excluded from the chance of unwanted interactions with other proteins under native conditions (Fig. 3B). The analysis recognizes the A strand and the CD and EF loops as the protein protective regions. Interestingly, the strong aggregation propensity of the naturally occurring truncated  $\beta 2m$  variant, devoid of the first six residues ( $\Delta N6$ )<sup>34</sup>, can be rationalized in the above context; in fact, by shortening the protective A strand, the C-terminal residues of the B strand, which are the most aggregation-prone, become solvent exposed. By the same token, one may also explain the behavior of D76N- $\beta 2m$ , the naturally occurring mutant responsible for a severe systemic amyloidosis<sup>10</sup>. In fact, the ‘fatal’ mutation leading to the most amyloidogenic isoform of  $\beta 2m$  so far described occurs in the protective EF loop.



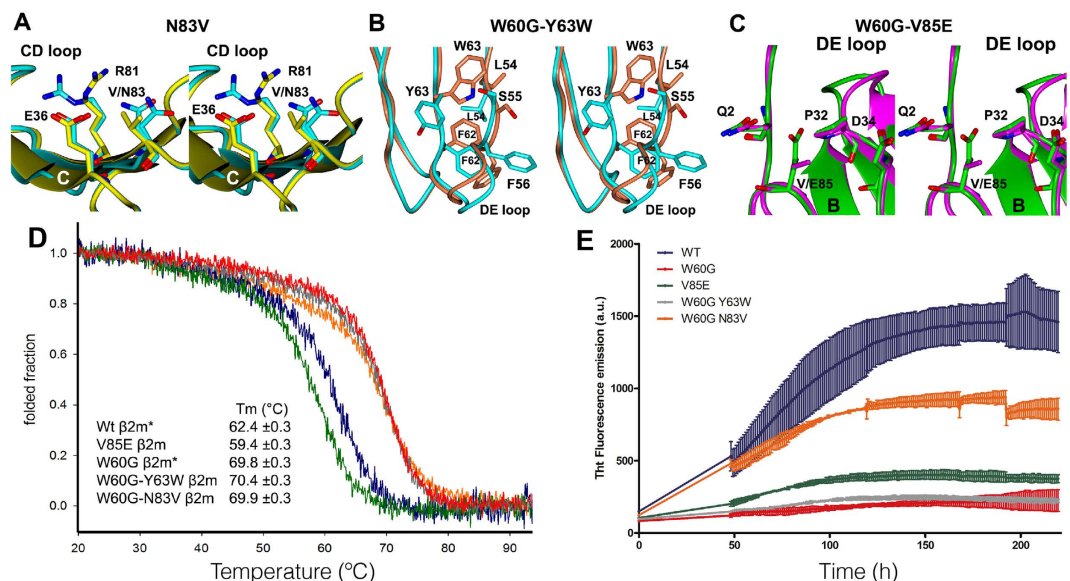


**Figure 4. Rational design of  $\beta 2m$  mutants.** (A) Aggregation-prone residues (intrinsic score  $> 1$ ) for which the protective effect of the W60G mutation is the highest, defined as those residues with the largest change in the structurally-corrected score between the wild-type (red) and W60G (black)  $\beta 2m$ . (B) Predicted aggregation propensity scores calculated using CamSol for all possible amino acid substitutions at the sites shown in panel A in the wild type (left box-plot in red) and in the W60G variant (right box-plot in black). The red and black dashed lines correspond to the aggregation propensity score of the WT and the W60G variant, respectively. Aggregation-protective and aggregation-promoting mutations selected for experimental validation are flagged. V85E and N83V mutations have been chosen as those with the largest predicted effect, while Y63W was chosen as a weakly aggregation promoting mutation. In each box the horizontal line denotes the median of the distribution; whiskers extend from the 5th to the 95th percentile of the distribution and boxes from the lower to the upper quartile.

The structurally corrected aggregation propensity profile for W60G shows an increase in the average protection of the aggregation-prone residues (Figs 3B,D and 4). Such an effect can be linked to the reduced strain at site 60 that propagates thorough the D strand in the 4-stranded  $\beta$ -sheet to the opposite  $\beta$ -sheet (Fig. 3D); indeed W60G shows on average a higher  $\beta$ -structure content distributed along all its  $\beta$ -strands (Fig. 2A–C) together with reduced fluctuations (Fig. S3).

**Rational design of mutants with different aggregation propensities.** The above analysis suggests then that the role of the W60G mutation in reducing aggregation is not exclusively due to stabilization of the protein. In particular, by optimizing the geometry of the D strand through the W60G mutation, all the aggregation-prone residues of the central  $\beta$ -strands, in both  $\beta$ -sheets, become on average more protected (Fig. 3). These observations suggest that by specifically modulating the average surface properties arising from the native state dynamics, it may be possible to tune the aggregation properties of  $\beta 2m$ .

To test the role of dynamics on  $\beta 2m$  native state aggregation we searched for mutations able to decrease the aggregation propensity of wild-type  $\beta 2m$ , on one hand, or increasing the aggregation propensity of W60G, on the other, without affecting their thermodynamic stabilities and structures. In Fig. 4A the residues with an intrinsic score higher than 1, for which the W60G mutation results in a large variation ( $> 0.2$ ) of the structurally corrected score with respect to the wild-type are highlighted. These seven residues are all located in the central strands of both  $\beta$ -sheets. Table S1 reports the amino acid substitutions, known within the  $\beta 2m$  protein family, that occur at such selected residue sites<sup>35</sup>; the sequence data highlight that none of these sites show strong residue conservation. Mutations of residues 65 and 67, in the E strand, have already been studied<sup>36</sup>, even though under different experimental conditions, showing that residues in this  $\beta$ -strand are critical for aggregation. Mutations in the F strand have not yet been investigated. By calculating the change in the protein intrinsic score upon all the possible single mutations at these seven sites (Fig. 4B), we suggest that the V85E  $\beta 2m$  variant should reduce the wild-type aggregation propensity, while two alternative specific mutations in the W60G  $\beta 2m$  variant (W60G-Y63W and W60G-N83V) are expected to weakly (the former) or largely (the latter) increase the W60G aggregation propensity. Indeed, the score of V85E  $\beta 2m$  is lower than that of W60G, the one of W60G-N83V is comparable to that of the wild-type while W60G-Y63W is slightly larger than that of W60G (Fig. 4B). All three mutations are located on the protein surface, and two of them are also naturally occurring in vertebrates other than human; hence they should cause little perturbation to the protein structure and stability. It is noteworthy that the effects of surface mutations introducing or removing surface charges, do not correlate simply with aggregation propensity: the V84D and Y86E mutations in acylphosphatase from *S. solfataricus* do not protect the protein from unfolding and aggregation<sup>37</sup>. Furthermore, a D to N systematic scanning in  $\beta 2m$  indicated that the removal of a negative charge triggers measurable effects in aggregation propensity only when the D to N mutation occurs at a specific site<sup>38</sup>.



**Figure 5. Structure, stability and aggregation of the β2m variants.** (A) The superimposed structures of the W60G and W60G-V83N mutants are shown in cyan and yellow, respectively. (B) Zoom of the (D and E) strands: the superposed structures of the W60G and W60G-Y63W are shown in cyan and coral, respectively. (C) Superimposed structures of wild-type β2m and the V85E variant, at the mutation site, shown in magenta and green, respectively. (D) Thermal unfolding of wild-type β2m, of the W60G variant and of the three surface mutants, monitored by far-UV CD at 202 nm. In the table the measured T<sub>m</sub> are reported (\*the T<sub>m</sub> of wild-type and W60G variants have been previously reported<sup>23</sup>). (E) Comparison of the kinetics of fibril formation of the various mutational variants of β2m analyzed in this work monitored by fluorescence in a thioflavin T assay.

**Structural characterization of three rationally-designed β2m surface mutants.** The three β2m surface mutants that we rationally designed (V85E, W60G-N83V, W60G-Y63W) were produced as recombinant proteins, purified to homogeneity, crystallized, and their structures determined at high resolution (1.75 Å, 1.49 Å, 1.70 Å, for V85E, W60G-Y63W, and W60G-N83V, respectively). Although the three crystals belong to the same space group, unit cell edges differ and crystal packing is distinct for each of the three mutants (see methods). The three structures display electron density of excellent quality along the polypeptide chains. Data collection and refinement statistics are reported in Table S2.

Inspection of the refined X-ray crystal structures shows that all three surface mutations are well tolerated in the β2m fold, and backbone conformational changes are relegated to loops (Table S3). Indeed, the largest structural deviations in the three mutants occur far from the mutation sites, at the AB loop (residues 12–21) that is typically in an open conformation in the crystal structures of isolated β2m, while it is closed over the rest of the β-sandwich when β2m is part of the MHC class I and in few structures of monomeric β2m<sup>39–41</sup>. Notably, the AB loop in a closed conformation is consistently observed in solution NMR structures<sup>42</sup>. In particular, for the crystal structures here reported of the W60G-Y63W and V85E mutants a closed AB loop conformation was observed; by contrast, the wild-type β2m, W60G and W60G-N83V 3D structures exhibit an open AB loop conformation. All these data suggest that the AB loop can easily adjust to different conformations, which, in the case of the present mutants, is likely the result of convenient molecular packing in the crystals.

Detailed inspection shows that the structure of the W60G-N83V mutant is perfectly superimposable onto that of W60G β2m. In fact, in W60G the side chain of N83, which is located on the F-strand, although surrounded by two charged residues (E36 and R81), does not establish any hydrogen bond with the neighboring residues; thus, the N83V substitution leaves both the conformation of residue 83 and of the surrounding side chains unaffected (Fig. 5A).

Residue 63 lies in the E strand, in close vicinity of the DE loop (residues 57–60). The Y63W mutation triggers just a minor reorientation of the DE loop backbone, but is responsible of extensive, though local, side chain reorganization. In the structure of the W60G-Y63W mutant the bulkier W63 side chain flips away from the DE loop and towards the D-strand, compared to what is observed in W60G. Such movement has two consequences: (1) the shift of S55, F56, L54 and F62 side chains away from W63, and (2) the DE loop moves about 2 Å towards residue 63 (Fig. 5B).

The comparison of the wild-type β2m and the V85E mutant structures reveals that the mutation does not elicit local rearrangements in the FG loop where E85 is located, nor in the surrounding residues. The 3D-structure of V85E mutant displays modified conformations in the DE and C'D loops; however, given that both regions are far from the mutation site, these differences may be due to the specific crystal packing (Fig. 5C).

**Stability of the rationally-designed β2m surface mutants.** The stability of all three mutants described above was evaluated by monitoring their thermal denaturation through far-UV CD measurements.

In keeping with the goals set (*i.e.* designing mutants that maintained the same thermodynamic stability either of wild-type  $\beta 2m$  or of the W60G mutant) and with the outcomes of the crystallographic analyses, none of the three mutations was found to alter dramatically the protein stability. Upon V85E mutation, the wild-type  $\beta 2m$  stability was slightly impaired, with a  $T_m$  value decreased by about 3 °C relative to wild-type  $\beta 2m$ . The negative charge of E85 lies 4.5 Å away from the carboxylate of D34, the electrostatic repulsion between these two residues may contribute to the slightly lower stability observed for the V85E mutant compared to the wild-type protein (Fig. 5C). Conversely, introduction of the Y63W and N83V mutations (added to the W60G mutation) did not vary significantly the  $T_m$  value recorded for the W60G mutant (Fig. 5D).

**Aggregation measurements of the rationally-designed  $\beta 2m$  surface mutants.** To further test the effects of the three designed mutations, the aggregation propensity of the V85E, W60G-Y63W, and W60G-N83V mutants was tested at pH 7.4 and in 20% TFE, conditions under which wild-type  $\beta 2m$  aggregates readily while the W60G mutant yields only small amounts of aggregates<sup>19,43</sup>. Figure 5E shows that the V85E mutation, despite being slightly less stable than the wild-type protein, almost totally abolished wild-type  $\beta 2m$  aggregation-propensity, and that the W60G-N83V variant presented a markedly increased aggregation trend with respect to W60G  $\beta 2m$ . The W60G-Y63W variant displayed the same very low aggregation propensity as the W60G mutant, suggesting that mutations should be chosen in order to maximize the difference of the total average CamSol score (Fig. 4B).

## Discussion

As protein aggregates are often thermodynamically stable in living organisms<sup>1</sup>, the avoidance of the aggregation process has been optimized through evolution by means of multiple strategies<sup>44</sup>. While efficient protection mechanisms work at the cellular level by regulating protein homeostasis<sup>45,46</sup>, many strategies to prevent aggregation concern the properties of the native states of proteins, including a high thermodynamic and kinetic stability, limited conformational fluctuations, and the presence of structural motifs that are incompatible with the cross- $\beta$  assembly<sup>1,4,47</sup>. As folded proteins are metastable with respect to aggregation<sup>3</sup>, mutation that destabilize the folded states can dramatically increase the rates of aggregation. Furthermore, mutations that change the accessibility to a high free energy state, without changing the overall stability of the protein native state, are enough to enhance aggregation<sup>11,48</sup>. Conversely, mutations that introduce structural protections on edge strands play a protective role independently of the overall protein stability<sup>37</sup>.

Here we investigated the different roles played by stability and dynamics in protein aggregation using human  $\beta 2m$ , a protein that, when released as a monomer and accumulated in patients' sera, may become the etiological agent of dialysis-related amyloidosis, forming amyloid deposits in joints, bones and muscles<sup>17</sup>. Esposito *et al.* reported previously an engineered  $\beta 2m$  variant bearing the W60G mutation, which was found to be thermodynamically very stable ( $C_m$  increased by 0.7 M GdHCl) and resistant to aggregation relative to the wild type  $\beta 2m$ <sup>19,20</sup>. To elucidate the molecular determinants underlying such distinct properties, we characterized the native state fluctuations of the wild-type protein and of the stable W60G mutant using NMR chemical shifts and RAM simulations, and analyzed the aggregation properties of the resulting conformational ensembles.

Our simulation results show that the W60G mutation greatly reduces the flexibility of the protein native state (Fig. S3), and increases the overall content of residues adopting  $\beta$ -structure (Fig. 2), in keeping with previous experimental data<sup>19,20,30</sup>. Moreover, our analysis shows that  $\beta 2m$  hosts three aggregation-prone regions located essentially in three  $\beta$ -strands that are central in the  $\beta 2m$  two  $\beta$ -sheets, *i.e.* strands B, E and F (Fig. 1A), and that in these regions those residues that are exposed on the surface are better protected on average in the W60G variant than in the wild-type (Fig. 3). The differences in the native state dynamics together with the similarity of their sequence-based intrinsic scores, explains the different behaviors observed under native and denaturing conditions, *i.e.* while both protein variants aggregate under denaturing conditions (pH 2.5), W60G shows little aggregation propensity under the standard aggregation protocol, at pH 7.4 and 20% TFE<sup>19,20</sup>.

To probe the role played by the residues in the three central  $\beta$ -strands in determining  $\beta 2m$  aggregation propensity, and to further investigate the protective effect of the W60G mutation, we designed three mutations aimed to modulate  $\beta 2m$  aggregation trends under native conditions, while leaving  $\beta 2m$  stability and structure unaltered. To this aim, we focused on mutations of protein surface residues that turned out to be better protected in the W60G ensemble compared to the wild-type protein. The V85E single site  $\beta 2m$  mutant, together with the W60G-Y63W and the W60G-N83V double mutants were accordingly designed and produced in recombinant form. Consistent with such rational approach, we found that all three surface mutants maintain unaltered the  $\beta 2m$  core structure (Fig. 5 and Table S2) and display fold stabilities very similar to wild-type  $\beta 2m$  (for V85E), or to W60G (for the two double mutants), respectively. Conversely, analysis of their aggregation properties indicates that surface mutations can indeed modulate  $\beta 2m$  aggregation propensity. The V85E mutation, on the F strand, abrogates completely wild-type  $\beta 2m$  amyloidogenicity, under the tested conditions, despite being slightly less stable than the wild-type; the W60G-N83V mutant (residue 83 falls on the F strand) gains a non-negligible aggregation propensity relative to W60G without any difference in protein stability. We cannot rule out that the fibrillar aggregate may form following a different aggregation pathway compared to wild-type  $\beta 2m$  as it was shown in details for mutants of acylphosphatase from *S. solfataricus*<sup>49</sup>. However in such case this phenomenon was coupled to very different thermodynamic stability of the native protein<sup>49</sup>. The W60G-Y63W double mutant, on the other hand, displays an aggregation trend comparable to W60G and in agreement with the small change in the aggregation score (Fig. 4B), indicating that this conservative mutation cannot encode noticeable changes of the aggregation propensity.

$\beta 2m$  hosts extensive aggregation regions on protein surface (Fig. 3), this being rooted in its sequence and structure that are specifically optimized to be part of the MHC class I. In particular, the aggregation-prone residues W60 and Y63 are buried at the interface between  $\beta 2m$  and the MHC class I heavy chain<sup>19</sup>, consistent with the observation that protein interface residues are typically aggregation prone<sup>50</sup>. On the contrary, the F

strand, which has not been previously studied, plays a pivotal role in the aggregation propensity of  $\beta 2m$ , and it is solvent-exposed in MHC class I. In particular, site 85 in human  $\beta 2m$  bears a dangerously aggregation-prone Val residue. Such apparent design error, however, could be reconciled with a recent observation<sup>15</sup> whereby the MHC class I heavy chain stabilizes markedly the  $\beta 2m$  fold.

In summary our results provide an explanation for the resistance to aggregation associated with the W60G mutation of  $\beta 2m$ . This mutation not only brings a significant increase in the thermodynamic stability of the protein, but also generates a more regular and less flexible overall structure in which the surface aggregation-prone regions are better protected. We suggest that such second feature is primarily responsible for the reduced aggregation propensity of this variant. Based on these observations, we showed that the aggregation propensity of  $\beta 2m$  can be tuned by mutating surface residues without affecting its structure and stability, thus highlighting the crucial role played by the dynamics of surface residues in protein aggregation. Our results elucidate a mechanism whereby the aggregation propensity of a protein can be modulated by mutations that change the average protection of its aggregation-prone surface residues with minor effects on the structure and stability of the protein itself.

## Materials and Methods

**Multidimensional NMR spectroscopy.** All data were recorded on a U-<sup>13</sup>C, U-<sup>15</sup>N labeled W60G  $\beta 2m$  sample 0.5 mM in phosphate buffer 70 mM, NaCl 100 mM, pH 6.6, at 310 K with a Bruker Avance 500 and with a Varian INOVA spectrometers operating at proton frequency of 500 and 800 MHz, respectively. Proton chemical shifts were referenced to 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt, DSS, whose resonance was set to 0.0 ppm. <sup>13</sup>C and <sup>15</sup>N chemical shifts were referenced indirectly to DSS, using absolute frequency ratio<sup>51</sup>. Sequence specific assignment were achieved using [1H, 15N]-HSQC and [1H, 13C]-HSQC together with 3D triple-resonance experiments, in particular by using the [15N, 1H]-HNCA/[15N, 1H]-HN(CO)CA for the identification of patterns of sequentially linked spin systems and 3D [15N, 1H]-HNCACB/[15N, 1H]-HN(CO)CACB pair of experiments, which allowed the identification of the amino acid types. 3D (H)CCH-TOCSY was used for the assignment of H $\alpha$  and H $\beta$  resonances.

The spectra were processed with Topspin 2.0 (Bruker Biospin) and analysed into the Sparky (T. D. Goddard and D. G. Kneller, SPARKY3, University of California, San Francisco) framework.

**Molecular dynamics simulations.** All the simulations in the present work were performed using GROMACS compiled with PLUMED and ALMOST. The system was simulated using the Amber03W force field in explicit TIP4P05 water. A time step of 2 fs was used together with LINCS constraints. van der Waals and short-range electrostatic interactions were cut-off at 0.9 nm while long range electrostatic was treated with the Particle Mesh Ewald method and a mesh size of 0.12 nm. The canonical ensemble was enforced by keeping the volume fixed and by thermostatting the system with the Nosé-Hoover thermostat at 310 K. The starting conformation for the wild-type was taken from the 1BMG X-Ray structure while for W60G was taken from the NMR structure 2VB5. The structures were protonated and solvated with 11500 water molecules in a dodecahedron box of 360 nm<sup>3</sup> of volume.

**Replica-averaged Metadynamics.** RAM simulations were performed using chemical shifts as replica-averaged restraints and bias-exchange metadynamics. The bias-exchange metadynamics approach combines replica exchange<sup>52</sup> with metadynamics, in which several metadynamics simulations are performed in parallel on different replicas of the system, each replica biasing a different collective variable (CV). Exchanges between the replicas are attempted periodically according to a replica-exchange scheme. Four replicas of the system were simulated in parallel at 310 K with a restraint applied on the average value of the CamShift back-calculated NMR chemical shifts where the force constant is set to 24 kJ/(mol ppm<sup>2</sup>).

Each of the four replica is bias along one of the following four collective variables, CVs: the anti-parallel  $\beta$  content (the ' $\beta$ ' CV), the parallel  $\beta$ -sheet content (the ' $\beta 3$ ' CV), the AlphaBeta collective variable defined over all the  $\chi 1$  angles for the hydrophobic side-chains (the 'AB' CV) and the AlphaBeta collective variable defined over all the  $\phi$  and  $\psi$  backbone dihedral angles of the protein (the 'bbAB' CV). The choice of the secondary structures CVs was guided by the fact that the protein is a full beta protein. The to AlphaBeta CVs are designed to sped up the fluctuations of the backbone and the side-chains irrelevant of their specific conformation. Gaussians deposition was performed with an initial rate of 0.125 kJ/mol/ps, a bias-factor of 8 and with  $\sigma$  values set to 0.1, 0.04, 0.16 and 0.25, for  $\beta$ ,  $\beta 3$ , AB and bbAB, respectively. Furthermore, in order to limit the extent of accessible space along each collective variable and correctly treat the problem of the borders, we set the bias as constant outside a defined interval for each CV; intervals were set to 19–29, 3–7, 14–30 and 47–71 for the four CVs, respectively. Each replica have been evolved for 200 ns, with exchange trials every 50 ps.

The sampling of the four replicas was used to generate a four-dimensional free energy landscape, as a function of the before mentioned CVs, where a set of microstates is identified by dividing the four-dimensional CV-space into a homogeneous grid of small dimensional hypercubes and their free energy is obtained using a standard weighted histogram analysis. Bi-dimensional FESes are obtained integrating out the variables not showed.

**Cloning, mutagenesis, expression and purification.** Recombinant vectors pEX containing genes encoding for W60G-Y63W, W60G-N83V variants, flanking NdeI/XhoI sites were purchased from Eurofins MWG Operon®. An empty vector pET-29b and the vectors carrying the  $\beta 2m$  variants were digested with NdeI and XhoI enzymes (New England Biolabs (UK) Ltd®) to obtain a linearized pET29 vector and the isolated genes. Wild-type  $\beta 2m$ , W60G-Y63W and W60G-N83V genes were inserted in pET-29b vector with a ligation reaction (Thermo Scientific Rapid DNA Ligation Kit, Life Technologies®). pET-29b-WT $\beta 2m$  plasmid was used as template DNA for Site-Directed Mutagenesis to obtain the V85E mutant. The following primers were used: 5'-GCC TGC



CGT GTG AAC CAT **GAG** ACT TTG TCA CAG CCC-3', 3'-GG GCT GTG ACA AAG **TCT** CAT GGT TCA CAC GGC AGG C-5'. All mutants were expressed and purified as previously reported<sup>34</sup>.

**Crystallization and X-ray structure determination.** All mutants were crystallized with sitting drops technique. Lyophilized W60G-Y63W, W60G-N83V and V85E  $\beta$ 2m mutants were solubilized in ddH<sub>2</sub>O at a concentration of 10 mg/mL, 8.5 mg/mL and 8 mg/mL respectively at 20 °C.  $\beta$ 2mW60G-Y63W crystals were obtained in 0.1 M HEPES pH 7.5, 20% (w/v) PEG 10K buffer (Crystal Screen HT<sup>TM</sup>, Hampton). The crystals were cryo-protected in 0.08 M HEPES pH 7.5, 18% PEG 10K, and 20% (v/v) glycerol and flash-frozen in liquid nitrogen. W60G-N83V crystallized in 0.1 M MES pH 6, 21% PEG 4K, 15% glycerol, 0.2 M ammonium acetate, while V85E mutant was crystallized in 0.1 M ammonium acetate pH 5.5, 30% PEG 4K, 15% glycerol and 0.2 M ammonium acetate. The crystals were flash-frozen using mother liquor as cryoprotectant.

X-ray data were collected at ESRF (European Synchrotron Radiation Facility of Grenoble–France) at the ID23-2 beam line. The diffraction data were analyzed and processed using MOSFLM<sup>53</sup>, the crystal symmetry was then verified by POINTLESS<sup>54</sup> and the intensities were merged with SCALA<sup>55</sup>. The crystal structure was determined by molecular replacement using MOLREP<sup>56</sup> using the structure of W60V  $\beta$ 2m mutant (pdb code: 2Z9T) as searching model<sup>22</sup>. The model molecules placed in the asymmetric unit were subjected firstly to a rigid-body refinement and then to a restrained refinement using Phenix Refine in the Phenix program suite<sup>57</sup>. Manual model building, addition of water molecules and ligands were then performed using the molecular graphic software Coot<sup>58</sup>.

All three mutants crystallized in C2 space group however the underlying crystal packing is different: the two double mutants (W60G-Y63W, W60G-N83V) have one molecule in the asymmetric unit (AU); however, the different conformation of the AB loop determines distinct intermolecular interactions in the crystals. Unusually, the AU of the V85E structure contains two molecules. These different crystal packings determine some changes in the protein regions involved in the intermolecular interactions, however, these differences are not deemed relevant for the discussion on the properties of the three mutants.

**Circular dichroism spectroscopy.** Thermal stability experiments, performed in the far-UV region, were carried out on a J-810 spectropolarimeter (JASCO Corp., Tokyo, Japan) equipped with a Peltier system for temperature control. The protein concentration was 0.1 mg/mL in 50 mM sodium phosphate pH 7.4. The temperature ramp measurements were recorded from 20 to 95 °C (temperature slope 50 °C/hour) in a 0.1 cm path length cuvette and monitored at 202 nm wavelength.  $T_m$  was calculated as the first-derivative minimum of the traces. Spectra before and after unfolding ramp were recorded (260–190 nm).

**Aggregation assays.** Samples (100  $\mu$ L) of recombinant variants W60G, V85E, W60G-Y63W, W60G-N83V, and wild-type  $\beta$ 2m at 100  $\mu$ M in 50 mM phosphate buffer (pH 7.4), 100 mM NaCl, 20% trifluoroethanol and containing 10  $\mu$ M Thioflavin T (ThT) (SIGMA), were incubated at 37 °C in Costar 96-well black-wall plates sealed with clear sealing film (4TITUDE) and were subjected to 900 rpm double-orbital shaking. In each well fragmented fibrils of wild type  $\beta$ 2m were added as seeds. Bottom fluorescence was recorded at 15-min intervals (BMG LABTECH FLUOstar Omega). Fluorescence was monitored in three or more replicate tests.

**NMR assignment deposition.** <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C assignments of W60G  $\beta$ 2m have been deposited to Biological Magnetic Resonance Bank with accession code 25809.

**Structure deposition.** Atomic coordinates and structure factors for the  $\beta$ 2m mutants W60G-Y63W, W60G-N83V, V85E have been deposited with the Protein Data Bank, with accession codes 5CFH, 5CKA and 5CKG, respectively.

## References

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890, doi: 10.1038/nature02261 (2003).
2. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15**, 384–396, doi: 10.1038/nrm3810 (2014).
3. Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J Am. Chem. Soc.* **133**, 14160–14163, doi: 10.1021/ja2017703 (2011).
4. Tartaglia, G. G., Pechmann, S., Dobson, C. M. & Vendruscolo, M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **32**, 204–206, doi: 10.1016/j.tibs.2007.03.005 (2007).
5. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805–808, doi: 10.1038/nature01891 (2003).
6. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395–1401, doi: 10.1039/b706784b (2008).
7. Chen, C. D. *et al.* Furin initiates gelsolin familial amyloidosis in the Golgi through a defect in Ca(2+) stabilization. *The EMBO J* **20**, 6277–6287, doi: 10.1093/emboj/20.22.6277 (2001).
8. Hammarstrom, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**, 713–716, doi: 10.1126/science.1079589 (2003).
9. Merlini, G. & Bellotti, V. Molecular mechanisms of amyloidosis. *N Engl J Med* **349**, 583–596 (2003).
10. Valleix, S. *et al.* Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. *N Engl J Med* **366**, 2276–2283, doi: 10.1056/NEJMoa1201356 (2012).
11. De Simone, A. *et al.* Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. *Proc. Natl Acad. Sci. USA* **108**, 21057–21062, doi: 10.1073/pnas.1112197108 (2011).
12. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
13. Achour, A. *et al.* Structural basis of the differential stability and receptor specificity of H-2Db in complex with murine versus human beta2-microglobulin. *J Mol. Biol.* **356**, 382–396, doi: 10.1016/j.jmb.2005.11.068 (2006).
14. Porcelli, S. A. & Modlin, R. L. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu Rev Immunol* **17**, 297–329 (1999).

15. Halabelian, L. *et al.* Class I Major Histocompatibility Complex, the Trojan Horse for Secretion of Amyloidogenic beta2-Microglobulin. *J Biol. Chem.* **289**, 3318–3327, doi: 10.1074/jbc.M113.524157 (2014).
16. Floege, J. & Ketteler, M. beta2-microglobulin-derived amyloidosis: an update. *Kidney Int Suppl* **78**, S164–171, doi: kid7823 (2001).
17. Gejyo, F. *et al.* A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Bioch. Biophys. Res. Comm.* **129**, 701–706 (1985).
18. Fogolari, F. *et al.* Molecular Dynamics Simulation Suggests Possible Interaction Patterns at Early Steps of {beta}2-Microglobulin Aggregation. *Biophys J* **92**, 1673–1681 (2007).
19. Esposito, G. *et al.* The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol. Biol.* **378**, 885–895 (2008).
20. Rennella, E. *et al.* Folding and fibrillogenesis: clues from beta2-microglobulin. *J Mol. Biol.* **401**, 286–297 (2010).
21. Gianni, S. *et al.* Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl Acad. Sci. USA* **111**, 14141–14146, doi: 10.1073/pnas.1405233111 (2014).
22. Ricagno, S., Raimondi, S., Giorgetti, S., Bellotti, V. & Bolognesi, M. Human beta-2 microglobulin W60V mutant structure: Implications for stability and amyloid aggregation. *Bioch. Biophys. Res. Comm.* **380**, 543–547 (2009).
23. Santambrogio, C. *et al.* DE-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci* **19**, 1386–1394 (2010).
24. Ami, D. *et al.* Structure, stability, and aggregation of beta-2 microglobulin mutants: insights from a Fourier transform infrared study in solution and in the crystalline state. *Biophys J* **102**, 1676–1684, doi: 10.1016/j.bpj.2012.02.045 (2012).
25. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-Averaged Metadynamics. *J. Chem. Theory Comput.* **9**, 5610–5617 (2013).
26. Camilloni, C., Robustelli, P., De Simone, A., Cavalli, A. & Vendruscolo, M. Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. *J Am. Chem. Soc.* **134**, 3968–3971, doi: 10.1021/ja210951z (2012).
27. Camilloni, C. & Vendruscolo, M. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J Am. Chem. Soc.* **136**, 8982–8991, doi: 10.1021/ja5027584 (2014).
28. Cavalli, A., Camilloni, C. & Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J Chem. Phys.* **138**, 094112 doi: 10.1063/1.4793625 (2013).
29. Katou, H. *et al.* The role of disulfide bond in the amyloidogenic state of beta(2)-microglobulin studied by heteronuclear NMR. *Protein Sci.* **11**, 2218–2229, doi: 10.1110/ps.0213202 (2002).
30. Gumral, D. *et al.* Reduction of conformational mobility and aggregation in W60G beta2-microglobulin: assessment by 15N NMR relaxation. *Magn. Res. Chem.: MRC* **51**, 795–807, doi: 10.1002/mrc.4018 (2013).
31. Piana, S. & Laio, A. A bias-exchange approach to protein folding. *J Phys. Chem. B* **111**, 4553–4559, doi: 10.1021/jp067873l (2007).
32. Sormanni, P., Aprile, F. A. & Vendruscolo, M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol. Biol.* **427**, 478–490, doi: 10.1016/j.jmb.2014.09.026 (2015).
33. Agostini, F., Vendruscolo, M. & Tartaglia, G. G. Sequence-based prediction of protein solubility. *J Mol. Biol.* **421**, 237–241, doi: 10.1016/j.jmb.2011.12.005 (2012).
34. Esposito, G. *et al.* Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci* **9**, 831–845 (2000).
35. Raimondi, S. *et al.* The two tryptophans of beta2-microglobulin have distinct roles in function and folding and might represent two independent responses to evolutionary pressure. *BMC Evol Biol* **11**, 159, doi: 10.1186/1471-2148-11-159 (2011).
36. Platt, G. W., Routledge, K. E., Homans, S. W. & Radford, S. E. Fibril growth kinetics reveal a region of beta2-microglobulin important for nucleation and elongation of aggregation. *J Mol. Biol.* **378**, 251–263 (2008).
37. de Rosa, M. *et al.* Edge strand engineering prevents native-like aggregation in *Sulfolobus solfataricus* acylphosphatase. *FEBS J*, doi: 10.1111/febs.12861 (2014).
38. de Rosa, M. *et al.* Decoding the Structural Bases of D76N ss2-Microglobulin High Amyloidogenicity through Crystallography and Asn-Scan Mutagenesis. *Plos One* **10**, e0144061, doi: 10.1371/journal.pone.0144061 (2015).
39. Azinas, S. *et al.* D-strand perturbation and amyloid propensity in beta-2 microglobulin. *FEBS J* **278**, 2349–2358, doi: 10.1111/j.1742-4658.2011.08157.x (2011).
40. Colombo, M., de Rosa, M., Bellotti, V., Ricagno, S. & Bolognesi, M. A recurrent D-strand association interface is observed in beta-2 microglobulin oligomers. *FEBS J* **279**, 1131–1143, doi: 10.1111/j.1742-4658.2012.08510.x (2012).
41. Eakin, C. M., Berman, A. J. & Miranker, A. D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat Struct Mol Biol* **13**, 202–208 (2006).
42. Verdone, G. *et al.* The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci* **11**, 487–499 (2002).
43. Natalello, A. *et al.* Wild type beta-2 microglobulin and DE loop mutants display a common fibrillar architecture. *Plos One* **10**, e0122449, doi: 10.1371/journal.pone.0122449 (2015).
44. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Bioch. Sci.* **24**, 329–332 (1999).
45. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11**, 777–788, doi: 10.1038/nrm2993 (2010).
46. Vendruscolo, M. Proteome folding and aggregation. *Curr Opin Struct Biol* **22**, 138–143, doi: 10.1016/j.sbi.2012.01.005 (2012).
47. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA* **99**, 2754–2759 (2002).
48. Neudecker, P. *et al.* Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366, doi: 10.1126/science.1214203 (2012).
49. Soldi, G., Bemporad, F. & Chiti, F. The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. *J Am. Chem. Soc.* **130**, 4295–4302, doi: 10.1021/ja076628s (2008).
50. Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. USA* **106**, 10159–10164, doi: 10.1073/pnas.0812414106 (2009).
51. Wishart, D. S. *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol. NMR* **6**, 135–140 (1995).
52. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
53. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ ESF-EACMB Newsletter on Protein Crystallography (1992).
54. Evans, P. Scaling and assessment of data quality. *Acta Cryst.* **62**, 72–82, doi: 10.1107/S0907444905036693 (2006).
55. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **50**, 760–763 (1994).
56. Vagin, A. A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J App. Crystallogr.* **30**, 1022–1025 (1997).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **66**, 213–221 doi: 10.1107/S0907444909052925 (2010).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **60**, 2126–2132 (2004).

## Acknowledgements

ESRF is kindly acknowledged for beamtime. This work was supported by the Italian Ministry of University and Research Project FIRB RBFR109EOS.

## Author Contributions

C.C., B.M.S., P.S., R.P., A.C., M.d.R., S.Z. and A.B. performed the experiments; C.C., A.C., M.B., M.V. and S.R. designed the study; C.C., G.E., M.B., V.B., M.V. and S.R. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Camilloni, C. *et al.* Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Sci. Rep.* **6**, 25559; doi: 10.1038/srep25559 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>